On the probabilistic modelling of the form function articulation for prosodic phenomena

Irina Nesterenko, Stéphane Rauzy, Daniel J. Hirst

Math. Sci. hum ~ Mathematics and Social Sciences (45e année, n° 180, *2007*(4), p. ??)

ON THE PROBABILISTIC MODELLING OF THE FORM ~ FUNCTION ARTICULATION FOR PROSODIC PHENOMENA

Irina NESTERENKO, Stéphane RAUZY, Daniel HIRST [1]

RÉSUMÉ – Modélisation probabiliste de l'interface « forme ~ fonction » pour des phénomènes intonatifs

Nous explorons l'application des méthodes hybrides, reposant à la fois sur des représentations symboliques (phonologiques) et probabilistes dans la modélisation de l'interface « forme ~ fonction » pour des phénomènes intonatifs. À partir d'une représentation symbolique ancrée dans l'acoustique du signal et en accord avec les principes de la phonologie intonative, deux modèles d'enchaînement des catégories tonales sont dérivées moyennant les méthodes de grammaires probabilistes. Deux modèles probabilistes sont testés : le modèle des bigrammes et le modèle des patrons ; leur performance est ensuite évaluée à l'aide de la mesure d'entropie. Ces modèles sont enfin testés en prédiction.

MOTS-CLÉS – Entropie, Grammaires probabilistes, Interface "Forme ~ Fonction", Modèle des patrons, Prosodie

SUMMARY – *Hybrid techniques based both on symbolic (phonological) representations and probabilistic information are applied in modelling the form ~ function interface for the prosodic phenomena. The symbolic representation is acoustically oriented and agrees with the principles of intonational phonology. On the basis of this representation, two models of tonal segments sequences are calculated according to the paradigm of probabilistic grammars. We test both the bi-grams model end the patterns model and the models' performance is further evaluated with the entropy measure. Finally, the probabilistic models are tested in prediction.*

KEYWORDS – Entropy, "Form ~ Function" Articulation, Patterns model, Probabilistic grammars, Prosody

INTRODUCTION

One of the central subjects of prosodic research is to specify and to model the ways in which prosody contributes to meaning. This process could receive the interpretation in terms of a mapping between formal and functional representations of prosodic phenomena. If such a perspective is adopted, it should be kept in mind that studying the relationship between prosodic forms and functions becomes rather circular if a clear distinction between the two is not made.

[1] Laboratoire Parole et Langage, Université de Provence, 29 av. R. Schuman 13621 Aix-en-Provence Cedex 1, France, {irina.nesterenko, daniel.hirst, stephane.rauzy}@lpl.univ-aix.fr

The main goal of the work we present is to propose a method to render explicit the form ~ function interface for intonational phenomena. We look as well for an algorithm for deriving the prosodic functions from the prosodic forms.

The prosodic model underlying our work is positioned within the autosegmental approach [Goldsmith, 1976] in its adaptation for intonation languages: the pitch curve is represented as a sequence of tones, which are completely separated from segments and syllables. Consequently, the tones constitute autonomous units and are treated as autosegments, which are associated with the segmental units but not part of them. In our study we seek to reveal the regularities in the ways the tones combine to signal discrete prosodic functions.

To do this, we have recourse to probabilistic methods quite in conformity with the probabilistic nature of cognitive processes and representations. Probabilistic methods were banished from linguistic research for years by the dominant generative grammar framework (cf. the discussion of the issue by Pierrehumbert [2001]). This current proned the study of linguistic competence and brought to the front the concept of grammaticality while the variability and variation issues were relegated to the domain of performance. By contrast, in probabilistic approaches, non-typical productions are not excluded from the analysis as not in agreement with the underlying grammar; rather they are treated as rare or marked structures. Probabilistic approaches are much more data-oriented and, hence, more in agreement with current tendencies in linguistics [Goldsmith, 2001].

The rest of this paper is organised as follows: we start with the presentation of the underlying prosodic model, which conditioned the initial data processing and information extraction. We present next the mathematical apparatus we had recourse to. In the data-oriented perspective, we further detail the annotations carried out for the training corpus. Section 4 presents the main findings of our study: the regularities captured for the sequences of tonal labels, the evaluation of probabilistic language models developed and the predictive power of these models. Finally we discuss the impact of the approach adopted and the future work which is required.

UNDERLYING PROSODIC MODEL

In this section we discuss the prosodic model underlying our empirical study and subsequent mathematical modelling. In our work prosody is interpreted as an organisational system [Beckman, 1996] that could be exhaustively specified via the analysis of tonal and rhythmical layers as well as that of prosodic phrasing [Di Cristo, 2000; Selkirk, 1995]. The representations developed for the three layers have recently been formalised in numerous phonological studies [Pierrehumbert, 1980; Ladd, 1992; Liberman, Prince, 1977; Nespor, Vogel, 1986]: highly developed formal representations of the prosodic phenomena were consequently proposed. To dispose of formal descriptions and representations is in our view a very important step for prosodic research since these representations allow us on the one hand to capture and model the phonetics ~ phonology interface, and on the other hand to investigate in a more formalised way how prosody contributes to the meaning.

The present study is limited to the analysis of the tonal component only, since it represents the quest for the methodology in the form ~ function articulation field. However it should be pointed out that a combined research of the tonal strata and of prosodic phrasing is necessary for the complete description of prosodic organisation.

Today, intonational phonology [Pierrehumbert, 1980; Pierrehumbert, Beckman, 1986; Ladd, 1996] is a reference framework in analysing and annotating tonal phenomena. It is an autosegmental-metrical approach [Goldsmith, 1976; Prince, Liberman, 1977]: the model has separate layers for representing segmental and tonal events and the tonal configurations are associated with metrically stressed syllables. The main principles of intonational phonology were incorporated into the ToBI prosodic annotation system [Beckman *et al.*, 2005] and since then, ToBI-inspired annotation systems have been proposed for many languages [Jun, 2005]. Notwithstanding, at the level of surface phonological representation the ToBI annotation system proceeds by melting together prosodic forms and prosodic functions (cf. the status of * or % symbols). Furthermore, though the developed representations postulate the existence of a special phonetic implementation module whose function is to translate the abstract phonological representation into f0 targets, in their internal logic, the ToBI inspired annotations try to reveal the mental constructs of the speaker-listener for the studied language. Consequently, the actual acoustic properties and configurations of fundamental frequency curve do not influence the choice of pitch accent label (cf. a body of studies on tonal alignment, particularly, [Arvaniti *et al.*, 2000]). Such a conception of association and alignment mechanisms crucially deviates from one developed by Bruce [1977] in one of the precursor works in the field of intonational phonology. It should be mentioned as well that a ToBI-inspired annotation does not exist for Russian, the language our study is based on.

Keeping in mind the reservations we have expressed above, in our study of the form ~ function interface for the Russian language, we turned to an alternative annotation scheme developed in the Laboratoire Parole et Langage of Aix-en-Provence: Momel-Intsint prosodic annotation protocol [Hirst, Di Cristo, 1998: Hirst *et al.*, 2000] which was recently coupled with the IF (for Intonation Functions) functional annotation system (Hirst 2005). The underlying prosodic model was conceived as a prosodic form ~ function interface donated with a multilevel architecture [Hirst *et al.*, 2000]. In this model, prosodic functions and prosodic forms are clearly separated and independent. This annotational scheme is founded on the assumption that in all languages a number of prosodic functions are expressed by a number of prosodic forms with the association between them being language-specific. At the same time, by postulating a multilevel organisation of the form ~ function interface, the authors insist on the fact that the formal aspects of prosody are not limited to the acoustics of the speech signal and they should not be directly related to the functional categories.

The prosodic model we refer to in our study comprises four distinct levels and four different representations. The analysis is tightly related to the acoustic signal, which appears at the bottom of the multilevel construct: the corresponding representation is one of the continuous fundamental frequency curve extracted from the signal by the appropriate software.

At the level of phonetic representation, two components are factored out from the f0 curve [Di Cristo, Hirst, 1986], a macroprosodic component and a microprosodic component. The first corresponds to a continuous smooth intonational curve, tightly associated with the prosodic meaning, while the second answers for the deviations from the smooth curve caused by the nature of the current segment. The macroprosodic component is further modelled via the application of the Momel algorithm [Hirst, Espesser, 1993]. This modelling is grounded on the definition of target points in time ~ frequency space: these target points correspond to the inflections in f0 curve where the slope is null (i.e. the first derivative equals zero). To obtain a smooth curve, the pitch

targets are linked by a quadratic spline function. Note that this conception of the target points has more in common with Bruce's conception [Bruce, 1977]. Another observation to keep in mind is that the level of phonetic representation is acoustically oriented. Note meanwhile that Vaissière [2002] views this model as production oriented: the tonal target points correspond to the sites where the speaker voluntary changes the direction of the fundamental frequency to achieve his/her communicational goals. The Momel algorithm is currently implemented under the Praat software [Boersma, Weenink, 2007].

At the upper, surface phonological level, the f0 targets receive a symbolic coding in terms of the Intsint prosodic alphabet [Hirst, Di Cristo, 1998]. This representation is further interpreted as a sequence of tonal segments in autosegmental framework. The Intsint alphabet comprises 8 distinct symbols. In this annotation the target points are characterised either globally with respect to the speaker's pitch range (via the long-term parameters of key and range; the corresponding labels are T(op), B(ottom) and M(edium)) or locally, by the reference to the preceding target (H(igher), L(ower), S(ame)). The H and L labels have the iterative variants D(ownstepped) and U(pstepped). The Intsint annotation scheme and the underlying prosodic model differ from the ToBI paradigm. We mentioned previously that the Momel-Intsint algorithm is more acoustically oriented, specifically in the definition of tonal targets. The issue of the alignment between tonal and segmental structures was not deeply investigated for it, though we think important to understand how the tonal segments are coordinated with the complex hierarchically organised prosodic constituency. Moreover, at the level of surface phonological representation, the tones are not organised between themselves, to form pitch accents or boundaries tones, tonetic primitives of the ToBI system.

The treatment of the articulation between the prosodic forms and functions is further reserved for the underlying phonological level. A set of prosodic functions could be specified (cf. [Di Cristo, 2000]): marking of the prosodic structure, speaker identification, turn-taking regulation or expression of emotions and affects. To represent the structural function of intonation, Hirst [1977, 2005] proposes an analysis in terms of two functional primitives such as *boundary* and *prominence*. These primitives can take the secondary attributes: it can be specified whether a prominence is [± nuclear] or [± emphatic]; on the other hand, the theories of juncture phenomena maintain the distinction between final and non-final boundaries ([± final] attribute). Such an annotation scheme exhibits the advantage of being independent from the analysis of prosodic forms.

The formal representations developed in the Momel-Intsint-IF paradigm are governed by one fundamental condition: this interpretability condition postulates that the representation at all intermediate levels must be interpretable at both adjacent levels, the more abstract and the more concrete. The interpretability condition follows from the considerations of the role of phonological representations: we assume that a phonological representation must provide the information necessary both for the pronunciation of the utterance and for its syntactic and semantic interpretation. In the present development of the prosodic model, the relation between formal and functional descriptions of prosodic phenomena has not been thoroughly investigated. Our study seeks to answer the question of the mapping mechanisms between prosodic functions and prosodic forms. Particularly, we explore the issue of whether there are any dependencies between Intsint labels in coding functional primitives as previously defined. To answer this question, we propose to investigate how prosodic labels of Intsint are distributed and how they combine in the set of prominence lending prosodic

forms. We particularly insist that the proposed description should be data-oriented and we found our study on a corpus of spontaneous speech, presented below.

1. EMPIRICAL STUDY DESIGN

1.1 CORPUS

Our study is based on a corpus of Russian spontaneous speech. This corpus was collected for the INTAS project 915 at the department of Phonetics, Saint-Petersburg State University. For the current study, the recordings of an informal spontaneous dialogue between two female speakers in their twenties were used and the productions of one of the speakers were analysed (17 minutes of speech including pauses).

The selected speech material was first processed with the Praat software to obtain Momel-Intsint annotation. This annotation crucially relies on two speaker dependent parameters: *key* and *range* (span), which define together the speaker's pitch range. Laver (1994) proposes five different readings and interpretations of the pitch range concept, which is also tightly related to the message informational organisation [Ladd, 1992; Brown, Yule, 1983]. Consequently, in order not to introduce a bias into the tonal annotation, the corpus was previously segmented into smaller units corresponding to one speaker's turn in the dialogue.

Besides, at the stage of target point detection (Momel algorithm), the detected points were manually corrected to achieve a perceptual equivalence between original and resynthesised versions (cf. IPO approach, [Hart *et al.*, 1990]) of the pitch contours. The corrected target points received an annotation in terms of the Intsint alphabet.

Given the Intsint annotation, next step is to extract the portions of the sequence of labels corresponding to the pitch accent bearing units (Prominence subset) and to the non-prominent units (Without Prominence subset). The extraction is done automatically with specially written Praat scripts. But first, the nature of the accent bearing unit should be clarified; our relevant choices are motivated in the next section.

1.2 THE MINIMAL DESCRIPTION UNIT

Given the Intsint annotation of the f0 curve, the next step is to extract the tonal configurations describing the prominence lending portions of the f0 curve. This extraction is done automatically by a Praat script, though first an additional annotation of the speech material is required: the corpus should be manually annotated in terms of prominence bearing units. In fact, if the syllable has the status of Tone Bearing Unit in autosegmental phonology, quite often a prominence lending pitch movement covers a span larger than a single syllable [Arvaniti *et al.*, 2000]. Moreover, the contrast between prominent and non-prominent units should be perceived.

The most general definition stipulates that an accented element together with the unaccented elements with which it contrasts, constitute an accentual unit. Following Garde [1968] the distinction is maintained between an accentable unit and an accentual unit: the former refers to the unit which bears the accent (usually the syllable) and the latter to the unit within which accentual contrasts are created (word, foot, accentual unit *etc.*). Yet the issue of the appropriate description unit, or domain within which the accentual contrast operates, is not a clear cut one in the prosodic studies and it is

intimately related to the established hierarchy of prosodic constituents for a given language in the underlying prosodic model.

Several other potential units have been proposed in the literature:

- prosodic word (or clitic group in Nespor & Vogel's (1986) nomenclature; not to be confused with a grammatical word defined by two spaces);
- foot (Abercrombie 1964), comprising one accented syllable and one or more unaccented syllables on its right or left side (cf. Hirst & Di Cristo (1998) for the discussion of the differences in metrical and tonal organisation between English and French);
- tonal association domain (Gussenhoven 1990, 2005), which is defined by the distribution of the pitch accents;
- tonal unit of the British school (Halliday 1967);
- accentual phrase (Jun & Fougeron 2000, Delais-Roussarie 2000).

Behind each candidate there is a developed theoretical basis, which we will omit here. Yet it should be mentioned, that the adequacy of each unit has never been tested empirically for one and the same language. It should be mentioned as well that the real domain on which the prominence lending movement is realised could have no phonological equivalent: Suomi's study of Finnish [Suomi, 2007] shows that from the phonetic perspective the tonal domain of accent corresponds to the first two morae, without any further reference to upper level units.

Moreover, numerous units from the list above could not be applied in the study of Russian language without substantial prior research. In fact, a very developed theory of prosodic domains and prosodic structure has been proposed for English; its transfer to other languages is less clear since terminological and conceptual differences make the interpretation of the findings difficult. For example, the foot is not frequently referred to in studies of rhythmical-metrical organisation of Russian; also, it couldn't be stated whether the Russian possesses the ïambic or trochaic foots. This is why at the present stage we decided to base our analysis on prosodic words.

So, the training corpus was annotated in terms of the prosodic word boundaries and further each prosodic word received its annotation according to whether the word bears or not a pitch prominence. We limited ourselves to one functional primitive, without any secondary attributes given the limited size of our corpus (we mentioned previously the methodological orientation of our study). Our corpus comprises 314 prosodic words associated with a pitch prominence and 511 non-prominent units. Further tests within the proposed paradigm will take into account a more developed functional annotation (note that both prominence lending pitch movement and boundary tone could be produced on the same prosodic word); these studies require much more annotated data.

The annotation done, we extracted two sets of tonal patterns and added two more symbols to the Intsint alphabet to clearly indicate the beginning and the end of the prosodic word-sized sequence. The distinction made between prominent and non-prominent units allowed us to calculate two probabilistic models of the tonal distributions and further compare their performance as to the automatic functional annotation based on the formal description.

1.3 MATHEMATICAL APPARATUS

In this section we will present the mathematical concepts and models we had recourse to in our work. Our objective was to answer the question how the Intsint tonal labels combine in the annotation of a prominence lending prosodic curve. Particularly, we search to reveal the regularities in the tonal patterns and consequently the dependency relations between the labels. As we mentioned previously, the apparatus of probabilistic grammars was applied: we looked for methods which would allow us to establish the probabilities over the tonal space, defined by the Intsint tonal categories. The analysis is founded on two mathematical concepts of interest: the concept of conditional probability and that of entropy.

Consider the general case when we dispose of N categories c_i to annotate a speech phenomenon. We assume as well to dispose of a training speech corpora, from which the distribution of tonal categories as well as their interdependencies are studied. Our task is then to estimate the probability of any produced sequence, say for example the time-ordered sequence (c_1, c_2, c_3, c_4) which corresponds to the f0 curve annotated in terms of Intsint tonal categories as (U, S, T, D). The probability of this sequence is calculated with the application of the concept of conditional probabilities:

$$P(c_1, c_2, c_3, c_4) = \pi_1 * \pi_2 * \pi_3 * \pi_4,$$

where $\pi_1 = P(c_1)$, $\pi_2 = P(c_2|\, c_1)$, $\pi_3 = P(c_3|\, c_1, c_2)$ and $\pi_4 = P(c_4|\, c_1, c_2, c_3)$. Herein the quantity $\pi_3 = P(c_3|\, c_1, c_2)$ stands for the conditional probability of the category c_3 given the preceding sequence of tonal labels (c_1, c_2). The formula above represents the general case though it could be simplified if we use a bigram model. Bigram models assume that only immediately adjacent tonal symbol influences the choice of the current one, i.e. the probability could be rewritten $P(c_1, c_2, c_3, c_4) = P(c_1) * P(c_2|\, c_1) * P(c_3|\, c_2) * P(c_4|\, c_3)$ under the bigram hypothesis. In our study both a bigram model and a Patterns model [Blache, Rauzy, 2006] were tested. The Patterns model is a new method belonging to the family of probabilistic finite state automaton approaches like n-gram models, for example (see [Rabiner, 1989] for a presentation of Hidden Markov Models). The Patterns model is characterized by an optimal extraction of the information content contained in the training corpora. Contrary to the n-gram model, the left context is not limited to a fixed number of symbols but rather takes into account the regularities observed in the corpus: if a sequence of tonal labels frequently occurs in the training database, the model calculates and memorises conditional probabilities for all the categories given the pattern.

To evaluate the performance of the model, we resort to the measure of entropy, which is the measure of the informational organisation of the system. For a tonal unit the entropy allows us to measure the informational charge of this unit and consequently to answer the question of how informative this unit is. Simultaneously, for a given distribution it quantifies the difference between this distribution and an equiprobable distribution of the categories. The entropy of the system varies between *0* and *lnN* (where *N* is the number of categories of the encoding scheme): an entropy of 0 characterises a completely deterministic system, while an entropy of lnN is found for the equiprobable distribution. We will introduce the concept of normalised entropy to bring the entropy values to the interval between 0 and 1. Subsequently, to evaluate the performance of the bigram model and patterns model we calculate the entropy of the probability distribution with and without the model.

When the probabilistic models of the label sequences were built we sought to test them in prediction, a step which will allow us to propose a methodology for the derivation of the prosodic functional primitives on the basis of the independent representation of prosodic forms. This task can be reformulated in mathematical terms as follows: given a prosodic word W with which is associated a tonal sequence $(c_1, \ldots, c_n)$ and given the probabilistic language models for prominent and non-prominent subsets, what is the probability for the considered unit to be perceived as bearing a pitch prominence? In other word, we evaluate the discriminative potential of the language probabilistic models. The evaluation is founded on the following principles: given the probabilistic models for prominent and non-prominent subsets, we calculate the probability of the tonal sequence in each of these models ($P_{WProminence}$ and $P_{WWithoutP}$). Two probabilities are calculated:

$$P_{W \in \ Prominence} = P_{WProminence} / (P_{WProminence} + P_{WWithoutP})$$

$$P_{W \in \ WithoutP} = P_{WWithoutP} / (P_{WProminence} + P_{WWithoutP}) = 1 - P_{W \in \ Prominence}$$

Next, a decision criterion is applied:

- if $P_{W \in \ Prominence} > 0.5$, the analysed prosodic word bears a pitch prominence with $P_{W \in \ Prominence}$ confidence level;
- if $P_{W \in \ Prominence} < 0.5$, the analysed prosodic word is non-prominent with a confidence level of $P_{WWithoutP}$.

The chosen decision criterion introduces the notion of confidence level, i.e. if the probability that the word belongs to the prominence category is of 0.99, we know that there is 1 chance in a hundred of making a mistake by attributing it to the prominence category.

The results of this prediction study were further evaluated with the measures of recall, precision, silence, noise and F-measure, traditionally used in information retrieval studies [Van Rijsbergen, 1979].

2. RESULTS

2.1 MODELS BUILDING STAGE

As stated previously, the goal of our study is to establish the facts about the distribution of Intsint tonal labels in the coding of the fundamental frequency curve associated with prominent and non-prominent accentual units.

To begin with, in the context of tune ~ text association problematics, we addressed the question of how often sequences of different sizes are found in the corpus (cf. Figure 1: the data are presented separately for prominent and non-prominent units). We observe that the longest string of tonal symbols comprises 6 symbols and is found only once in the corpus. A more important observation concerns the fact that already at this stage we can observe the difference between prominence lending and non-prominence lending curves: for the prominent accentual units, 78.4 % of the curves are modelled with two or three symbols, while 73 % of non-prominent units are associated with no more than 2 symbols. Such a result is consistent with the modelling of the intonational contour between the functional intonational events in terms of transitions.

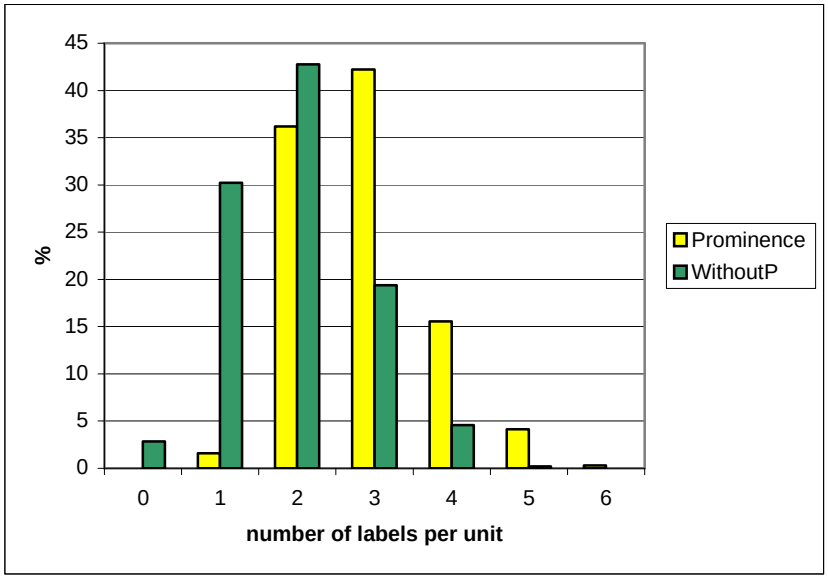

FIGURE 1. Frequency distribution of sequences of different sizes

If we analyse the frequency distribution of the Intsint tonal labels separately for prominent and non-prominent subsets of our corpus, we could state that the tonal labels associated with the speaker/utterance pitch range as well as the non-iterative tones L and H are more frequent in the conding of the prominence lending pitch movements. These differences are more important for T and H labels (0.07 *versus* 0.008 for T, 0.063 *versus* 0.019 for H). No crucial differences are observed for iterative tones U and D. If we evaluate the observed differences with the entropy measure (evaluating so the information weight of these frequency data), for the prominence bearing units we obtain a value of normalised entropy of 0.933, close to 1, which points at the quasi-equiprobable distribution. These data are to be contrasted with the facts stipulated for ToBI pitch accent labels: as Taylor [2000] points out, the distribution of ToBI labels is very uneven, 80% of the accents being annotated as H* . We might mention once again that Momel-Intsint algorithm provides a purely formal coding of the fundamental frequency curve.

For the non-prominent units, the normalised entropy is 0.852, which indicates that the frequency distribution of the tonal labels bears more information load.

Next, four probabilistic models founded on the conditional probabilities were calculated: Bigram and Patterns model separately for prominent and non-prominent subsets. The evaluation data are summarised in the Table 1.

		Entropy	Normalised entropy	mutual information
bigram model	Prominent units subset	1.389	0.603	0.759
	Non-prominent units subset	1.168	0.508	0.792
patterns model	Prominent units subset	1.265	0.549	0.883
	Non-prominent units subset	1.136	0.493	0.825

TABLE 1. Entropy and normalised entropy for four probabilistic models built

From the data of the table 1, we can see that to take into account one of the probabilistic models of the frequency distributions for the Intsint tonal labels reduces considerably the normalised entropy of the model. The conditioning effect is more marked for the non-prominent units (i.e. smaller values of the normalised entropy). At the same time, if we compare the performances of the bigram and Patterns models, the effect of the more complex model is more marked for the prominent units. Note that for this quantitative result we can propose a linguistic explanation as well: it has always been stated that the pitch movements associated with functional primitives are more complex than the transitions realised on non-prominent units. Equally, we have seen that the non-prominence lending pitch movements were more frequently modelled with no more than two Intsint symbols; consequently, we couldn't expect a great impact of the Patterns model.

2.2 PREDICTION STUDY

The confusion matrix and the corresponding evaluation statistics for the bi-gram model are presented in Tables 2 and 3. The recall quantifies the accuracy of the model, i.e. the proportion of correctly predicted cases over the data, the silence being its complementary measure: the overall accuracy of the model is 0.79. On the other hand, the precision measures the proportion of correctly predicted cases over all the cases placed in the same category by the model, the noise measure communicating the proportion of erroneously classified observations. Consequently, for the optimal performance of the model, the couple < recall, precision > should show maximal values: the f-measure statistics allow to take into account simultaneously the values of precision and recall. The bi-grams model benefits from an f-measure of 78%: we can conclude that there is plenty of room for the improvement, though the overall quality of the prediction is rather good.

	Predicted no-prominence	Predicted prominence	TOTAL
No-prominence	437	74	511
Prominence	94	220	314
TOTAL	531	294	657

TABLE 2. Confusion matrix for bigram model

Measure	
Precision	0.77
Recall	0.79
Noise	0.23
Silence	0.21
F-Measure	0.78

TABLE 3. Evaluation statistics

We present in Figure 2 the error level as a function of confidence level: these data allow us to evaluate the reliability of the corresponding functional annotation. For example, if we consider the cases with a confidence level greater than 0.9, we know that the classification could be erroneous in 5 cases out of one hundred. Though, for the cases with the confidence level less than 0.7, the probability of mistake is around 44%. If the proposed probabilistic model is implemented in a prosodic annotator, the proposed diagnostics tool could be used for an additional marking of the unsure cases.

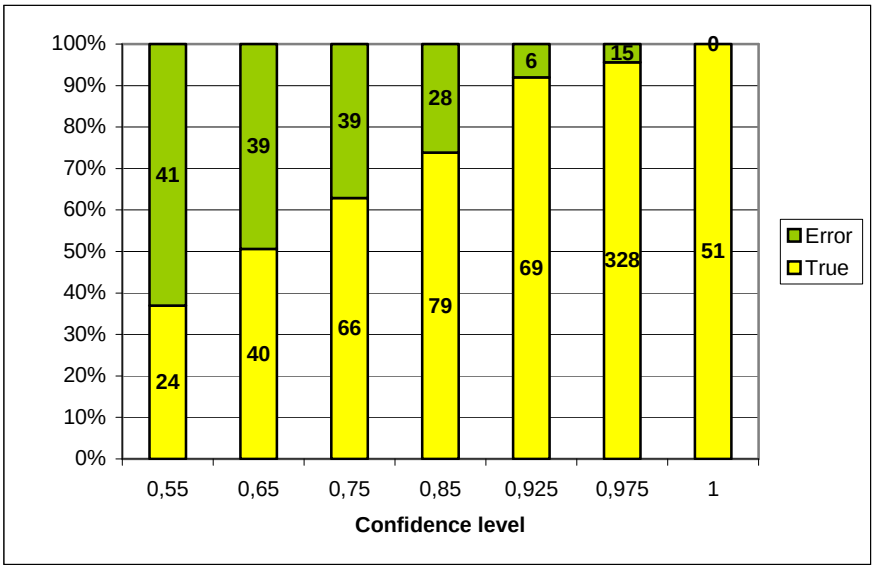

FIGURE 2. Error rate as a function of confidence level for bigram model

The same procedure was applied to evaluate the predictive potential of the Patterns model (the confusion matrix and evaluation statistics from Tables 4 and 5, confidence level data are summarised in Figure 3). We note a slight improvement of the model's performance, which seems to be essentially related to the prominence category (as we mentioned previously, the bigram model provides a sufficient modelling for the subset of non-prominent units).

	Predicted no-prominence	Predicted prominence	TOTAL
No-prominence	419	92	511
Prominence	71	243	314
TOTAL	490	335	661

TABLE 4. Confusion matrix for Patterns model

Measure	
Precision	0.8
Recall	0.79
Noise	0.2
Silence	0.21

F-Measure	0.79

TABLE 5. Evaluation measures

In general, the prediction study seems to indicate interesting paths for applications of the probabilistic models of language in speech technology research. The methodology we propose deduces the independently defined prosodic functions on the basis of prosodic forms representations and modelling. Moreover, we introduce the measure of confidence level, on the basis of which a special marking could be provided if the label assignment is less reliable.

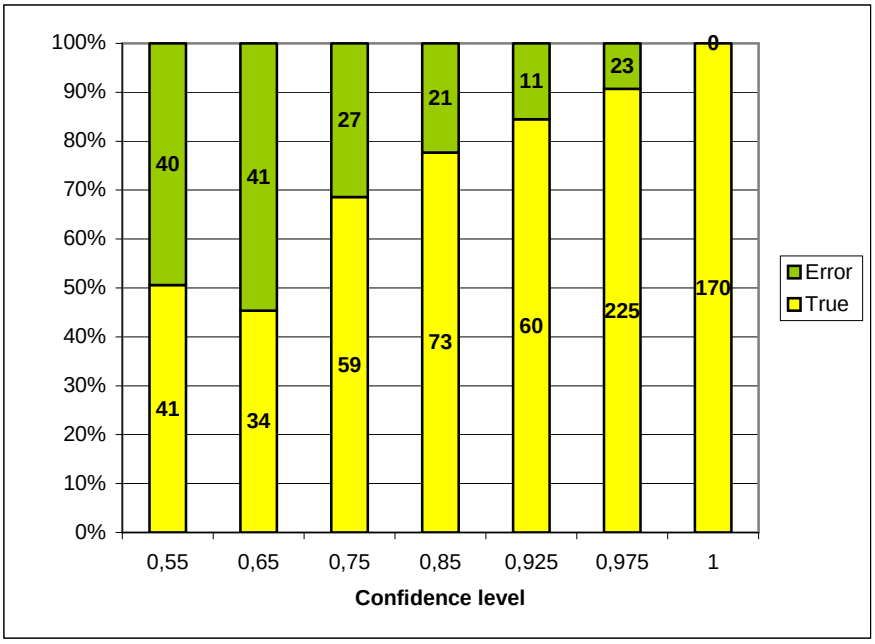

FIGURE 3. Error rate as a function of confidence level for Patterns model

3. CONCLUSIONS

Methodologically oriented, the present work sketches an approach for prosodic information retrieval, based on both symbolic and probabilistic information. The algorithm was tested on a corpus of Russian spontaneous speech provided with the Intsint formal prosodic annotation and with a rudimentary functional annotation, these annotations being subject to an independency constraint. Only one functional primitive without any secondary specifications was considered, i.e. melodic prominence, for which the distributions of the Intsint tonal labels were modelled.

Our starting assumption was that there are dependency relationships between the Intsint tonal labels, which form the patterns associated with prominence lending prosodic curves. Indeed, the probabilistic models allowed us to reveal these relationships. What is more, we note that while the pitch curves associated with non-prominent prosodic words are satisfactorily modelled by a bigram model, for the complex prominence lending pitch curves, a better fit was achieved with a Patterns model. This observation points out that the model for prominent units functions at least as a second order Markov model; though a more accurate model requires that the left

context be of variable size. On the basis of the analysis realised we can state that combining symbolic and probabilistic methods allows to capture generalisations about the form ~ function articulation for prosodic phenomena.

The probabilistic models of tone sequences could be further incorporated into the predictive heuristics for automatic corpus annotations in terms of prosodic functions. We propose to introduce in such applications the concept of confidence level to explicitly model the reliability of the proposed annotations; we can imagine special perceptual studies designed for the less reliable fragments. As we have particularly insisted on the interpretability condition imposed on the representations developed in the underlying prosodic model, the probabilistic models could as well be integrated into speech synthesis applications, in which prosodic forms are derived from prosodic function. Besides, the probabilities integrated allow us to keep track of the variability and contextual variation proper to human discourse and so to enrich the rather prescriptive algorithms presently at use in synthesis systems.

Given the size of the test corpus, we limited the present study to only one functional category. Simultaneously, the tonal component of the prosody was studied separately, without any reference to the hierarchical prosodic structure. The described paradigm allow us to incorporate the hypothesis that more detailed representation of prosodic organisation could influence the choice of tonal labels. To test this hypothesis, more of appropriately annotated data are needed. Another fruitful path for further research is opened with the cross-linguistic perspective in order to capture language specific features in the mapping between prosodic forms and prosodic functions.

BIBLIOGRAPHY

ABERCROMBIE D., *Elements of General Phonetics*, Edinburgh University Press, 1967.

ARVANITI A., LADD D.R., MENNEN I., "What is a starred tone? Evidence from Greek", M. Broe, J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, Cambridge, Cambridge University Press, 2000, p. 119-131.

BECKMAN M.E., "The parsing of prosody", *Language and Cognitive Processes* 11, 1996, p. 17-67.

BECKMAN M.E.. PIERREHUMBERT J.B., "Intonational Structure in Japanese and English", *Phonology Yearbook* 3, 1986, p. 255-309.

BECKMAN M.E., HIRSCHBERG J., SHATTUCK-HUFNAGEL S., "The original ToBI system and the evolution of the ToBI framework", S.-A. Jun (ed*.) Prosodic Typology – The Phonology of Intonation and Phrasing*, 2005.

BLACHE P., RAUZY S., "Mécanismes de contrôle pour l'analyse en Grammaires de Propriétés", *Proceedings of the Conference Traitement Automatique des Langues Naturelles* (TALN), Leuven, Belgium, April 10-13, 2006, p. 415-424.

BOERSMA P., WEENINK D., "Praat : a system for doing phonetics by computer", 1995-2007. Available from <http://www.fon.hum.uva.nl/praat/>

BROWN G., YULE G., *Discourse analysis*, Cambridge University Press, 1983.

BRUCE G., "Swedish Word Accents in Sentence Perspective*", Travaux de l'Institut de Linguistique de Lund* 12, Institut de Linguistique de Lund, Lund, Suède, 1977.

DELAIS-ROUSSARIE E., "Vers une nouvelle approche de la structure prosodique", *Langue Française* 126, 2000, p. 92-112.

DI CRISTO A., “Interpréter la prosodie”, *Actes des XXIII^e Journées d’Étude sur la Parole*, Aussois, 13-23 juin 2000, p. 13-29.

DI CRISTO A., HIRST D., “Modelling French micromelody: analysis and synthesis”, *Phonetica* 43(1), 1986, p. 11-30.

GARDE P., *L'accent*, Collection SUP, Le Linguiste 5, Paris, Presse Universitaires de France, 1968.

GOLDSMITH J., “An overview of autosegmental phonology”, *Linguistic Analysis* 2, 1976, p. 23-68.

GOLDSMITH J., “The unsupervised learning of natural language morphology”, *Computational Linguistics* 27, 2001, p. 153-198.

GUSSENHOVEN C., “Tonal association domains and the prosodic hierarchy in English”, S. M. Ramsaran (ed.), *Studies in the pronunciation of English. A commemorative volume in honour of A.C. Gimson*, London, Routledge, 1990, p. 27-37.

GUSSENHOVEN C., “Transcription of Dutch Intonation”, Jun S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford, Oxford University Press, 2005, p. 118-145.

HALLIDAY M.A.K., *Intonation and grammar in British English,* The Hague, Mouton, 1967.

HART J., COLLIER R., COHEN A., *A Perceptual Study of Intonation. An Experimental Phonetic Approach to Speech Melody, Collection Cambridge Studies in Speech Sciences and Communication,* Cambridge (UK), Cambridge University Press, 1990.

HIRST D., *Intonative Features. A Syntactic Approach to English Intonation*, Collection Janua Linguarum, Series Minor 139, La Haye, Mouton & Co, 1977.

HIRST D., “Form and function in the representation of speech prosody*”, Speech Communication* (Special Issue) 46(3-4), 2005, p. 334-347.

HIRST D., DI CRISTO A., *Intonation Systems. A Survey of Twenty Languages*, Cambridge (UK), Cambridge University Press, 1998.

HIRST D., DI CRISTO A., ESPESSER R., “Levels of description and levels of representation in the analysis of intonation”, M. Horne (ed.), *Prosody: Theory and Experiment*, Dordrecht, Kluwer Academic Press, 2000, p. 51-87.

HIRST D., ESPESSER R., “Automatic modelling of fundamental frequency using a quadratic spline function”, *Travaux de l'Institut de Phonétique d'Aix* 15, 1993, p. 71-85.

JUN S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press, 2005.

JUN S.-A., FOUGERON C., “A phonological model of French intonation”, Botinis A. (ed.), *Intonation : Analysis, Modelling and Technology*, Dordrecht, Kluwer Academic Publishers, 2000, p. 209-242.

LADD D.R., “An Introduction to Intonational Phonology (Prosody)”, Docherty G. J., Ladd D. R. (eds), *Papers In Laboratory Phonology II. Gesture, Segment, Prosody,* 1992, p. 321-334.

LADD D.R., *Intonational Phonology*, Collection Cambridge Studies in Linguistics 79, Cambridge (UK), Cambridge University Press, 1996.

LAVER J., *Principles of Phonetics*, Collection Cambridge Textbooks in Linguistics, Cambridge (UK), Cambridge University Press, 1994.

LIBERMAN M., PRINCE A.S., “On Stress and Linguistic Rhythm”, Goldsmith J.A. (ed.), *Phonological Theory. The Essential Readings,* USA, Blackwell Publ., 1977, p. 392-404.

NESPOR M., VOGEL I., *Prosodic Phonology*, Dordrecht, Foris Publication, 1986.

PIERREHUMBERT J., “*The phonology and phonetics of English intonation*”, PhD dissertation, MIT, 1980.

PIERREHUMBERT J., *Stochastic phonology*. *GLOT* 6, 2001, p. 1-13.

RABINER L.R., “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, *Proceedings of the IEEE 77*, 1989, p. 257-286.

SELKIRK E., “Sentence prosody: intonation, stress and phrasing”, J. Goldsmith (ed.), *Handbook of Phonological theory*, Cambridge, Basil Blackwell, 1995, p. 550-569.

SUOMI K., “On the tonal and temporal domains of accent in Finnish*”, Journal of Phonetics* 35(1), 2007, p. 40-55.

TAYLOR P., “Analysis and synthesis of intonation using the tilt model*”, Journal of the Acoustical Society of America* 107(3), 2000, p. 1697-1714.

VAISSIÈRE J., “Cross-linguistic prosodic transcription”, N.B. Volskaja, P.A. Skrelin, N.D. Svetozarova (eds), *Problems and methods in experimental phonetics*, St.-Petersburg State University, 2002, p. 147-164.

VAN RIJSBERGEN C.J., *Information Retrieval,* 2nd edition, Glasgow, University of Glasgow, 1979.